APPROXIMATE COUNTING: AN ALTERNATIVE APPROACH

Peter Kirschenhofer and Helmut Prodinger

Department of Algebra and Discrete Mathematics
Technical University of Vienna, Austria

Abstract. In this note an alternative analysis of approximate counting is presented by using a lemma from the calculus of finite differences instead of the Mellin integral transform.

R. Morris [6] has proposed a probabilistic algorithm that maintains an approximate count in the interval 1 to n using only about $\log_2 \log_2 n$ bits. Approximate counting (with a binary base) starts with counter $C = 1$. After n increments C should contain a good approximation to $\lfloor \log_2 n \rfloor$; thus C should be increased by 1 after another n increments approximately. Since only C is known the algorithm has to base its decision on the content of C alone. The following principle is used to increment the counter:

$$C := C + \begin{cases} 0 \text{ with probability } 1 - 2^{-C} \\ 1 \text{ with probability } 2^{-C} \end{cases}$$

Compare Flajolet [2] for a more detailed description. In the same paper a detailed analysis of this algorithm is presented and the following results are shown:

Theorem 1. *After n succesive increments the average content $\overline{C_n}$ of the counter satisfies:*

$$\overline{C_n} \sim \log_2 n + \frac{\gamma}{\log 2} - \alpha + \frac{1}{2} + \delta_1(\log_2 n) \ ;$$

the variance σ_n^2 satisfies

$$\sigma_n^2 \sim \sigma_\infty^2 + \delta_2(\log_2 n)$$

where

$$\sigma_\infty^2 \sim \frac{\pi^2}{6 \log^2 2} - \alpha - \beta + \frac{1}{12} - \frac{1}{\log 2} \sum_{k \ge 1} \frac{1}{k \sinh(\theta k)}$$

with

$$\alpha = \sum_{k \ge 1} \frac{1}{2^k - 1} \ , \ \beta = \sum_{k \ge 1} \frac{1}{(2^k - 1)^2} \ , \ \theta = \frac{2\pi^2}{\log 2} \ ,$$

We thank an anonymous referee for his/her guidance concerning a better presentation of this paper.

and $\delta_1(x)$ as well as $\delta_2(x)$ are periodic functions with period 1, mean value 0, and amplitude less than 10^{-4}.

Flajolet achieves this result by computing the probability $p_{n,l}$ of having counter value l after n increments:

$$p_{n,l} = \sum_{j=0}^{l-1} \frac{(-1)^j \, 2^{-\binom{j}{2}} \, (1 - 2^{-(l-j)})^n}{Q_j \, Q_{l-1-j}} \; , \tag{1}$$

where

$$Q_n = (1 - \frac{1}{2})(1 - \frac{1}{4}) \dots (1 - \frac{1}{2^n}). \tag{2}$$

The quantities

$$\overline{C_n} = \sum_{l=1}^{n} l \, p_{n,l} \text{ and } \sigma_n^2 = \sum_{l=1}^{n} l^2 \, p_{n,l} - (\overline{C_n})^2 \tag{3}$$

are then analyzed using real analysis and certain properties of the Mellin integral transform.

In the sequel we analyze the quantities in (3) by an alternative approach which avoids completely the unpleasant real analysis by making use of a classical lemma from the calculus of finite differences. This method has been attributed by Knuth [5] to S. O. Rice.

Lemma 2. *Let C be a curve surrounding the points $1, 2, ..., n$ in the complex plane and let $f(z)$ be analytic inside C. Then*

$$\sum_{k=1}^{n} \binom{n}{k} (-1)^k f(k) = -\frac{1}{2\pi i} \int_C < n; z > f(z) dz,$$

where

$$< n; z >= \frac{(-1)^{n-1} n!}{z(z-1)...(z-n)}.$$

Moving the contour of integration it turns out that the asymptotic expansion of the alternating sum is obtained via

$$\sum \operatorname{Res} \left(< n; z > f(z)\right) \tag{4}$$

where the sum is taken over all poles different from $1, ..., n$. (Flajolet and Sedgewick have used this technique to simplify the analysis of digital search trees in [3].)

In order to apply Lemma 2 we write (1) as

$$P_{n,l} = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \sum_{j=0}^{l-1} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j \, Q_{l-1-j}} \, 2^{-(l-j)k}. \tag{5}$$

By Euler's *partition identities* [1]

$$\sum_{j \ge 0} \frac{(-1)^j 2^{-\binom{j}{2}} x^j}{Q_j} = \prod_{m \ge 0} \left(1 - \frac{x}{2^m}\right)$$

and

$$\sum_{j\ge 0}\frac{x^j}{Q_j}=\prod_{m\ge 0}\Big(1-\frac{x}{2^m}\Big)^{-1},$$

the inner sum in (5) is the coefficient

$$[x^{l-1}]\sum_{j\ge 0}\frac{(-1)^j 2^{-\binom{j}{2}}x^j}{Q_j}\sum_{s\ge 0}2^{-(s+1)k}\frac{x^s}{Q_s}$$

$$=2^{-k}[x^{l-1}]\prod_{m=0}^{k-1}\Big(1-\frac{x}{2^m}\Big).$$

Thus

$$\overline{C_n}=\sum_{l=1}^{n} l\,p_{n,l}=\sum_{k=0}^{n}\binom{n}{k}(-1)^k 2^{-k}\sum_{l=1}^{n} l\,[x^{l-1}]\prod_{m=0}^{k-1}\Big(1-\frac{x}{2^m}\Big),$$

where the summation on l can be extended to infinity without changing the value. Since

$$\sum_{l\ge 1} l\,[x^{l-1}]\,g(x)=g'(1)+g(1)$$

we get immediately the following identity for $\overline{C_n}$ which we state as a Proposition, since it is interesting in its own right.

Proposition 3.

$$\overline{C_n}=1-\sum_{k=1}^{n}\binom{n}{k}(-1)^k 2^{-k}Q_{k-1}. \tag{6}$$

With

$$Q_z=\frac{\prod_{j\ge 1}\big(1-\frac{1}{2^j}\big)}{\prod_{j\ge 1}\big(1-\frac{1}{2^{z+j}}\big)}$$

we may write (6) as

$$\overline{C_n}=1-\sum_{k=1}^{n}\binom{n}{k}(-1)^k f(k)\quad\text{with}\quad f(z)=2^{-z}\,Q_{z-1}. \tag{7}$$

The most significant pole of $< n; z > f(z)$ different from $1, ..., n$ is $z=0$, and the residue is

$$-\frac{H_n}{L}+\frac{1}{2}+\alpha\sim-\log_2 n-\frac{\gamma}{L}+\frac{1}{2}+\alpha,$$

with $L=\log 2$ and H_n the n-th Harmonic number. Further poles in $z=2k\pi i/L$ with residue equal to

$$\frac{1}{L}\Gamma\left(\frac{2k\pi i}{L}\right)$$

give rise to the periodic fluctuation $\delta_1(x)$.

The next significant poles have real part $z=-1$ (simple!), so that the error term is of order n^{-1},

A similar approach allows us to evaluate the variance σ_n^2 in a few lines. Since

$$\sum_{l\ge 1} l^2\,[x^{l-1}]\,g(x) = g''(1)+3g'(1)+g(1)$$

we have

$$\sum_{l=1}^{n} l^2\,p_{n,l} = 1+\sum_{k=1}^{n}\binom{n}{k}(-1)^k f(k)$$

with

$$f(k)=2^{-k}Q_{k-1}(2T_{k-1}-3),\quad T_{k-1}=\sum_{i=1}^{k-1}\frac{1}{2^i-1}.$$

The continuation of Q_k has been referred above; T_k can be continued via

$$T_z=\alpha-\sum_{i\ge 1}\frac{1}{2^{z+i}-1}.$$

The most important pole is again $z=0$ (triple) with residue

$$\frac{2}{L^2}\left(\frac{H_n^2+H_n^{(2)}}{2}+H_n\,L\,(\frac{1}{2}-\alpha)\right)+\alpha^2-4\alpha-\beta-\frac{2}{3}$$

$$\sim \log_2^2 n+(\log_2 n)(\frac{2\gamma}{L}-2\alpha+1)+\alpha^2-2\alpha-\beta$$
$$+\frac{\gamma}{L}+\frac{\gamma^2}{L^2}-\frac{2\alpha\gamma}{L}+\frac{\pi^2}{6L^2}-\frac{2}{3}.$$

Subtracting $\overline{C_n^2}$ yields for the variance (apart from smaller order terms)

$$\frac{\pi^2}{6L^2}-\alpha-\beta+\frac{1}{12}-\frac{2}{L^2}\sum_{k\ge 1}\left|\Gamma\left(\frac{2k\pi i}{L}\right)\right|^2+\delta_2(\log_2 n), \tag{8}$$

where the series originates from the mean of $\delta_1^2(x)$. Observe that

$$|\Gamma(iy)|^2=\frac{\pi}{y\,\sinh \pi y},\ y\in\mathbf{R},$$

to obtain the form of σ_∞^2, whence Theorem 1.

Finally, it is striking to note that the constant σ_∞^2, whose numerical value is

$$\sigma_\infty^2=0.76301418710991095777\ldots$$

differs from the simpler

$$\frac{1}{2\log 2}+\frac{1}{24}=0.76301418711114837034\ldots$$

only in the eleventh digit after the decimal point.

Indeed, using a transformation result for Dedekind's η -function

$$\eta(\tau) = e^{\pi i \tau/12} \prod_{n\geq 1} (1 - e^{2\pi i n \tau}), \quad \Im \tau > 0\,,$$

the constant σ_∞^2 appearing in (8) turns out to be

$$\frac{1}{2L} + \frac{1}{24} - \frac{2}{L} h_1(\frac{2\pi^2}{L}) + \frac{2\pi^2}{L^2} h_2(\frac{4\pi^2}{L}),$$

where

$$h_1(x) = \sum_{k\geq 1} \frac{(-1)^{k-1}}{k(e^{kx}-1)} \text{ and } h_2(x) = \sum_{k\geq 1} \frac{e^{kx}}{(e^{kx}-1)^2},$$

so that the numerical coincidence follows by the smallness of the involved functions $h_1(x)$ and $h_2(x)$.

Compare [4], where a closely related constant with relevance to digital search trees was evaluated by the authors and J. Schoissengeier.

Finally it should be noted that the extension to arbitrary bases of the counter can be analyzed in the same style.

References

1. G. E. Andrews, *The Theory of Partitions*, Addison Wesley, 1976.
2. P. Flajolet, *Approximate Counting: A Detailed Analysis*, BIT **25** (1985), 113–134.
3. P. Flajolet and R. Sedgewick, *Digital Search Trees Revisited*, SIAM J.Comput. **15** (1986), 748–767.
4. P. Kirschenhofer , H. Prodinger, and J. Schoissengeier, *Zur Auswertung gewisser numerischer Reihen mit Hilfe modularer Funktionen*, Zahlentheoretische Analysis II (E. Hlawka, ed.), Lecture Notes in Math. 1262, Springer, Berlin, 1987, pp. 108–110.
5. D. E. Knuth, *The Art of Computer Programming, Vol. 3*, Addison-Wesley, 1973.
6. R. Morris, *Counting Large Numbers of Events in Small Registers*, Comm.ACM **21** (1978), 840–842.

TU Vienna, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria